THE EXPERIMENTAL DETERMINATION OF STANDARDS IN FIRST-YEAR ALGEBRA

H. O. RUGG
School of Education, University of Chicago

A report to the Mathematics Section of the Illinois High-School Conference of a preliminary investigation for the purpose of: (1) establishing certain methods by which standards for measuring the outcomes of a year's instruction in high-school algebra may be constructed; (2) stating tentative results obtained in eight Illinois high schools; (3) making certain criticisms of the learning and teaching process in algebra.

The conduct of this investigation is based upon the following statement of the ultimate aim of first-year algebra: The successful use of algebraic symbolism in meeting new problematic situations, either, (1) purely algebraic situations; (2) general mathematical situations (such as are found in other mathematics courses other than algebra); (3) non-symbolic situations including various types of "practical" or "applied" problems not expressed in mathematical language.

I. THE OUTCOMES OF FIRST-YEAR ALGEBRA

Believing that some statement of the assumed outcomes of first-year algebra should be made the basis for further analysis, the following tentative statement of these outcomes is made:

A. *Immediate, specific, and preparatory outcomes.*—These include the comprehension, interpretation, and manipulation of the specific "mechanical" operations involved in algebraic solutions; e.g., the four fundamentals as used with various type problems, factoring, removal of parentheses, etc. Certain of these are to be used in many specific algebra problems and in the solution of other mathematical problems; they are to be mastered as tools, preparatory to the taking of other mathematical courses, as well as for use as automatic tools for the solution of all types of "applied" problems. These involve primarily the learning of rules—the formation of specific habits of manipulation.

B. *Immediate generalized outcomes.*—These involve recourse to selective, analytic, and conceptualizing abilities; ability to apply principles, in addition to ability to remember rules and to make certain fundamental habitual adjustments in the solution of general practical and applied problems and in solution of problems belonging primarily to other mathematical fields.

C. *Remote and less tangible outcomes.*—The development of the ability to deal with general number concepts and "think quantitatively"; the development of attitudes of (1) orientation in algebraic or general mathematical fields containing problematic situations; (2) confidence in one's ability to use successfully algebraic symbols in meeting new situations; (3) a broadened intellectual background or perspective for the general cultural comprehension and interpretation of the scientific methods by which technical problems may be solved, i.e., the development of a "scientific attitude." (In the preliminary attack on the general problem of standards in this field it is not proposed to consider this third group of outcomes. We believe standards in first-year algebra should deal primarily with the "mechanical" processes—i.e., those which should be made automatic—and only very little with these other more indefinite ones.)

2. METHOD OF PROCEDURE

At the 1913 meeting of the Mathematics Section of the Illinois High-School Conference a committee was instructed to begin the study of establishing standards by which the results of one year's high-school algebra might be measured. The problem, as attacked by the chairman of that committee (Mr. R. L. Modesitt of the Eastern Illinois Normal School), consisted of an attempt to determine the relative degree of difficulty inhering in a set of "mixed" algebra problems (i.e., a list containing both "mechanical" and "applied" problems), as shown by the varying proportions of a group of fifty pupils who were asked to solve the problems. This attempt, resulting in an adaptation of Thorndike's "mixed scale" (discussed below), was to be extended during the past school year. Other duties compelling Mr. Modesitt to give up the work, the conduct of the investigation was turned over to the present writer. Mr. E. L. Mayo kindly consented to assist in the work but has been

prevented from much active co-operation by unexpected and uncontrollable circumstances arising in connection with the administration of the classes in his school. He and his colleagues have helped, however, by giving the Thorndike tests in their classes. Thus the writer is alone responsible for the assumptions, line of attack, and present analysis and conclusions from the data.

It should be emphasized in starting that we regard this investigation to date as of a preliminary nature only and not as one which has enabled us to state definite and final standards. The problem being turned over to the writer so late in the year caused the collection of data and organization of the work to be somewhat hurried. The material presented in this report is therefore more open to criticism of the "conditions" under which the data were secured than will be the case in the extension of the investigation during the coming year. We consider, however, that a sound beginning has been made on a very complex problem, the conclusion of which can come only in the course of years of careful and co-operative experimentation.

The procedure since April, 1915, may be outlined as follows:

1. Preliminary statement of the aims and the outcomes of the teaching of high-school algebra.

2. Determination of basic method of designing and constructing tests for measuring efficiency in first-year algebra. (The investigation of the validity of the assumptions underlying Thorndike's "mixed" scale, concerning (*a*) the relative difficulty of problems and (*b*) the validity of testing pupils with a scale composed of both "mechanical" and applied problems.)

3. Classification of the subject-matter of first-year algebra and the determination of a list of specific operations whose efficiency should be tested.

4. Determination of principles which should govern the selection of test problems.

5. Selection of problems composing each test with preliminary individual experimentation to standardize the arrangement and timing of these problems.

6. Final organization of test sheets and giving of tests in eight school systems in Illinois according to standard directions.

7. The correction of the test papers; original tabulation of scores obtained; statistical treatment of results to state typical conditions; a minute analysis of the specific errors made in each problem.

8. The interpretation of tabular and statistical material and conclusions concerning the progress of the experiment to date.

3. GENERAL CLASSIFICATION OF SUBJECT-MATTER AND LIST OF SPECIFIC OPERATIONS TO BE MASTERED

The subject-matter in first-year algebra is first roughly classified to fit the assumed outcomes as follows:

A. *All material of a mechanical nature*, necessitating complete automatism—the establishment of definite groups of habits through continued drill. The writer assumes that the following list of operations should be absolutely mastered as a result of the drill given in first-year algebra (there is stated in parentheses after each point the number of the test in which the efficiency of this operation is measured):

1. Addition and subtraction of positive and negative numbers.
2. Law of signs in multiplication (Tests 1 and 2).
3. Law of signs in division (omitted in design of tests).
4. Addition and subtraction of monomials (Test 2; Test B–3).
5. Addition of polynomials (omitted in design of tests).
6. Removal of parentheses (Tests 1 and 2).
7. Multiplication of monomials (Tests 1 and 2; Test B–17).
8. Multiplication of polynomials (Tests 1 and 2).
9. Multiplication of monomials by monomials, polynomials by polynomials, and monomials by polynomials (Tests 1 and 2).
10. Division of monomials, reduction of quotients (Test 3, problems 7, 14, 21, 28).
11. Use of exponents (Test 3).
    a) without parentheses:
    A. Multiplication of monomials involving
        1) integral numerical exponents (problems 1, 3, 8, 10, 15, 16, 17, 19, 22, 23, 24).
    b) with parentheses (problems 4, 11, 18, 25, 27).
    c) negative integral exponents (problems 5 and 12).
        2) literal exponents (problems 2, 9, 16, 17).
        3) fractional exponents (problems 6, 13, 20, 26).
    B. Division of monomials involving exponents (problems 7, 14, 21, 28).

12. Division of polynomials (omitted in design of tests).
13. Squaring binomials (omitted in design of tests).
14. Substitution (Test 5; Test B, problems 1 and 18).
15. Factoring monomials (Test 4, problems 2, 7, 12, 17, 22).
16. Factoring of difference of two squares (Test 4, problems 1, 6, 11, 16, 21).
17. Factoring trinomial squares (Test 4, problems, 4, 9, 14, 19, 24).
18. Factoring trinomials of form ax^2+bx+c (Test 8, problems 3, 8, 12, 16, 20).
19. Solution of equations of first degree with one unknown (Test 6; Test B–14).
20. Solution of fractional equations (Test 7).
21. Solution of quadratic equations (Test 8).

B. *All material of an original nature* which, though necessitating constant use of definitely learned systems of habits, primarily involves independent thinking in applying fixed methods to new problematic situations:

1. Applied problems leading to equations of first degree with one unknown (Test B, problems 4, 5, 9, 10, 11, 12, 16, 19, 21, 23, 25).
2. Applied problems involving ratio and proportion (Test B, problems 8, 13, 15, 22).
3. Applied motion problems (Test B, problems 16 and 17).
4. Applied problems involving translation of verbal expression into algebraic symbolism (nearly all of problems in Test B).

It should be emphasized again that our attention has been centered primarily on the organization of the automatic operations and that the list given here of the types of verbal problems is by no means considered exhaustive. Furthermore in connection with such topics as: the graph and its application, the formula, solution of equations in two unknowns, groups of three or more linear equations, etc., we are far from a final decision as to their place in the course of study. It will be the aim of next year's study to establish a standard in this field.

4. THE DETERMINATION OF A VALID METHOD FOR MEASURING EFFICIENCY IN THE LISTED OPERATIONS AND TYPES OF PROBLEMS

The design of tests to measure efficiency in algebra implies a definite and detailed analysis of the specific operations and processes to be mastered in the study of the subject. The design to be outlined here is based on the analysis made above.

A. *Types of design.*—There are at least three types of design.

Type I. The Mixed Scale: Professor E. L. Thorndike assumes that one mixed test, containing "mechanical" problems of many types and "original" problems involving many sorts of fundamental operations, will be a valid measure of efficiency in elementary algebra.

Type II: On the other hand, the opposite extreme view concerning algebra tests would have each specific operation tested separately; i.e., such elemental processes as addition of monomials and polynomials and the other fundamentals would be measured by separate tests.

Type III: Between these two extremes it is possible to occupy a middle ground, by building a test-system which will test the larger and more important operations in a reasonably detailed manner, at the same time combining certain of the most elemental operations in one test. The first type of design obviously cannot measure in any valid way many of the more important outcomes of the study of algebra. The second method, if carried to its logical extreme, would build up a test-system so detailed and cumbersome that no school system could hope to use it in the periodic determination of pupil efficiency. Granted that the ideal "measuring-stick" would measure specifically every operation used in algebraic solution, classroom and administrative requirements would prohibit its use. To get a usable measure, therefore, we must compromise in our desire for detailed analysis. The writer has imposed the criterion that the giving of the completed algebra standards to any class shall occupy not more than two or three class periods of 45 minutes each.

B. *Fundamental criteria for the design of an algebra scale.*—To measure algebraic abilities we must have a basis for determining the *relative difficulty* of problems and tests. There are four possible criteria: (I) the teacher-judgment basis; (II) "proportion-of-pupils-solving" basis; (III) the quantitative enumeration and objective analysis of the steps and operations involved in various types of problems; (IV) the solution of all types of tests by individuals under carefully controlled conditions (i.e., individual laboratory tests) with detailed individual discussion and comments on particular problems and difficulties. The teacher-judgment

method assumes that the relative difficulty of algebraic problems and processes can be determined by the uncontrolled judgment of teachers of algebra. The second method assumes that the relative difficulty of problems will vary directly with the proportion of a large number of pupils able to solve the problems in question correctly.

I. The Teacher-Judgment Basis: Implicit in the design of all of Professor E. L. Thorndike's "scales" (handwriting, drawing, composition, and algebra) is the assumption that "relative difficulty" or "relative merit" will vary as the proportion of "expert" judges varies who rank certain samples of student work as better or poorer than other samples. In order to apply the method to algebra he submitted 25 problems of various types (printed herewith in the order of their final ranking) to 200 teachers of algebra, asking them to "rank" the problems in order of difficulty. From the returns he selected a "scale" of nine problems, some of the automatic type and some of the applied type, some simple and some very complex, by taking the problems D, K, A, T, H, E, I, V, W, which an approximately equal percentage of his group of judges had ranked as of successively greater degree of difficulty. That is, approximately 80 per cent of 200 judges rank K more difficult than D; A more difficult than K, etc., throughout the list. Thorndike's deduction is, after some statistical manipulation, that the interval of difficulty between each two consecutive problems, D, K, A, etc., is equal and that these problems taken together represent a scale by which we may measure the efficiency of 20 weeks' instruction in elementary algebra. (It should be said that Professor Thorndike states that a better scale could be designed by using other methods. We are criticizing here Thorndike's acceptance and continued use of this method, *not* his failure to recognize the greater validity of other methods.)

LIST OF PROBLEMS USED BY PROFESSOR THORNDIKE IN DESIGNING ALGEBRA SCALE

We state with each problem the percentage of pupils who solved the problem correctly

(97.6) D. If $a=4$ and $b=2$ what does $a+b$ equal? Answer..............

(97.0) K. If $a=4$ and $b=0$ what does $a+b$ equal? Answer..............

(91.7) A. If $x+3a=5a$ what does x equal? Answer....................

(87.0) L. If $3x+4=2x+8$ what does x equal? Answer..................

(64.5) O. If $a=3$ and $b=2$ what does a^2-ab equal? Answer............

(62.8) S. If $a=6$ and $b=1$ what does $2ab-ab^2$ equal? Answer..........

(36.1) T. Find the average midnight temperature for the week in which the daily midnight temperatures were 15, 3, 0, -7, -9, 6, and 17 degrees. Answer..

(53.2) P. If $x-2a+b=2x+2b-4a$ what does x equal? Answer.........

(61.0) U. If $\frac{x}{a+b}=a-b$ what does x equal? Answer..................

(34.3) G. If $a=6$ and $b=3$ what does $\sqrt{a}\sqrt{2b}$ equal? Answer..........

(53.9) B. The circumference of a circle $=2\pi r$. $\pi=3\frac{1}{7}$. $r=$ the length of the radius of the circle in question. If the diameter of a bicycle wheel is 28 inches, how many inches is the circumference? Answer...

(19.0) C. If $\frac{6x+7}{5}-\frac{2x-1}{10}=4\frac{1}{2}$ what does x equal? Answer.............

(62.8) Q. If $\frac{4}{x+2}+\frac{7}{x+3}=\frac{37}{x^2+5x+6}$ what does x equal? Answer........

(17.9) H. If $\frac{1}{a}-\frac{1}{x}=\frac{1}{x}-\frac{1}{b}$ what does x equal? Answer..................

(8.3) E. If $2+\frac{\frac{x}{a}-1}{\frac{2}{a}}=0$ what does x equal? Answer..................

(0.0) I. A man has a hours to spend riding with a friend. How far can they ride together, going out at the rate of b miles an hour and just covering the return trip at the rate of c miles an hour? Answer...

(3.5) R. Let e stand for the safe load that can be hoisted by a hemp rope. Let c stand for the circumference of a rope. If $l=100\ c^2$ for any hemp rope, how many pounds are a safe load for a hemp rope $2\frac{1}{4}$ inches in circumference? Answer..........................

M. If $\frac{x+a}{x-a}-\frac{x-a}{x+a}-\frac{x^2}{a^2-x^2}=1$ what does x equal? Answer........

X. At what time between 6 and 6:30 o'clock are the hands of a watch at right angles to each other? Answer......................

Y. If $x=\frac{a+b}{2}$ what does $\left(\frac{x-a}{x-b}\right)^3-\frac{x-2a+b}{x+a-2b}$ equal? Answer......

J. If $\frac{a+b}{b+c}=\frac{c+d}{d+a}$ prove that $a=c$, or that $a+b+c+d=0$.

N. There are two thermometers or scales to measure temperature. The Fahrenheit scale (F.) is the one we commonly use. The other is called the Centigrade scale (C.). A temperature of 32 degrees on the F. scale=0 degrees on the C. scale; 33.8 degrees on the F. scale=1 degree on the C. scale; 35.6 degrees on F.=2 degrees on C.; 50 degrees on F.=10 degrees on C.; 14 degrees on F.= −10 degrees on C.

a) What on the C. scale=70 on the F. scale? Answer.....

b) What on the C. scale=4 degrees *below* zero on the F. scale? Answer..................................

c) What on the F. scale=20 degrees on the C. scale? Answer..

V. How much water must be added to a pint of "alcohol, 95 per cent pure" to make a solution of "alcohol, 40 per cent pure"? Answer..

F. A cube containing eight cubic inches was plated with copper. The difference in the weights of the cube before and after the plating was 0.139 lb. One cubic inch of copper weighs 0.315 lb. Form an equation from which the approximate thickness of the copper plating could be calculated. State whether the approximate estimated thickness by your equation would be less or more than the exact thickness.

W. Given that $2x-3$ is less than $x+5$ and that $11+2x$ is less than $3x+5$, to find the limits (i.e., the values) between which x lies.

(It is understood that the pupil has not had any special training in inequalities or limits. This problem is, so to speak, an original exercise.)

II. The Proportion-of-Pupils-Solving Basis: The foregoing basis (I) being open to criticism on analytical and psychological grounds, the writer reports herewith a slight beginning that he has made in investigating the validity of the method. The progress made has consisted of a comparison of Thorndike's results with those obtained by having students solve the same list of problems. His list of problems has been worked by 169 students in the Joliet Township High School, 92 of whom had just finished first-year algebra, 58 of whom had finished it one year before taking the tests, and 19 of whom had studied it two or more years before. The problems were arranged on the sheet in the exact order of difficulty as determined by Thorndike's rank method, giving what little influence the time element might exert to the support of his ranking. Table I below summarizes the number and percentage of this group

TABLE I

PERCENTAGE OF 169 HIGH-SCHOOL PUPILS SOLVING CORRECTLY THE FIRST 19 OF THORNDIKE'S TEST PROBLEMS

Problems are lettered

	Problems	D	K	A	L	O	S	T	P	U	G	B	C	Q	H	E	I	R	M	X
Number	92 pupils	89	87	85	82	57	49	39	56	61	34	47	12	66	22	11	0	2	0	0
	58 pupils	57	58	54	51	40	24	17	27	34	17	31	15	36	6	3	0	2	0	0
	19 pupils	19	19	16	14	12	5	6	7	8	7	13	5	4	2	0	0	2	0	0
	Total, 169 pupils	165	164	155	147	109	78	62	90	103	58	91	32	106	30	14	0	6	0	0
Percentage	92 pupils	96.7	94.6	92.4	89.2	61.9	53.3	42.4	60.9	66.3	37.0	51.1	13.0	71.8	24.0	12.0	0	2.3	0	0
	58 pupils	98.3	100	93.1	88.0	70.0	41.4	30.0	46.6	58.6	30.0	53.4	25.9	62.1	10.4	5.2	0	4.0	0	0
	Total, 169 pupils	97.6	97.0	91.7	87.0	64.5	46.1	36.7	53.2	61.0	34.3	53.9	19.0	62.8	17.8	8.3	0	3.5	0	0

of students who solved each of the first 19 problems correctly. (They were given one class period of 45 minutes to solve the entire list of 25 problems. All attempted to solve more than the first 19 and it is assumed here that the pupils had sufficient time to satisfy themselves that they could or could not work these 19 problems. That is, we believe that the element of "pressure" or "drive" due to the "time" factor does not operate here.)

What does the tabulation in Table I show? It shows that, instead of being solved successfully by gradually decreasing groups of pupils, the problems fall into a few sharply differentiated groups. Problems D, K, A, L, (Group A) can be solved by nine-tenths or more of the pupils who have had one year's algebra instruction; problems O, U, Q and possibly P and B seem to be about equally difficult, in general being solved by about two-thirds as many pupils as the problems of the first group; T and G are in another class of "difficulty" and C and H are in still another; E, I, R, M, X are again in a class by themselves, being so difficult to solve that practically no pupils can work them. Our tabulation confirms Thorndike only in the fact that the *order* of problems is the same, the interval of difficulty being decidedly unlike as determined by the two methods. To bring out this point more clearly note the size of the "interval of difficulty" between the seven problems on his scale and that determined approximately by the other basis.

TABLE II

COMPARISON OF RELATIVE DIFFERENCE IN DIFFICULTY BY TWO METHODS

Relative Difference in Difficulty (Thorndike)	Problems	Percentage of Group Solving	Approximate Relative Difference of Difficulty Based on Proportion of Pupils Solving
1 diff.	D	97.6	1 diff.
2 "	K	97.0	1 "
3 "	A	91.7	2 "
4 "	T	36.1	7 "
5 "	H	17.9	9 "
6 "	E	8.3	10 "
7 "	I	0.0	11 "

One outstanding fact appears: The judgment of teachers cannot be taken as a safe criterion for estimating the relative frequency

with which pupils can be expected to solve various types of problems. The writer would go further and say that the judgments of teachers cannot be taken as a safe criterion in determining the relative difficulty of algebra problems, and that of the two methods thus far discussed the "proportion-of-pupils-solving" method is the more valid. Any method of determining difficulty of problem solution must be based on a sound and minute analysis of subject-matter and of the psychological processes involved in the solution. A most elementary analysis would reveal at once the distinction in difficulty-to-the-student, between the large group of "mechanical" operations which have been made more or less automatic by drill and the group of applied or original problems in which little or no drill has been given. That this analysis is necessary is shown by our tabulation, viz.: Practically all pupils can solve the simple automatic operations of substitution when expressed in "drill" form as in problems D, K, A; practically no pupils can use exactly the same processes when expressed in "original" form as in R. Approximately half of the students can solve fairly complex equations of the first or second degree in one unknown when expressed in familiar "drill" form; practically no pupils can successfully use the same processes when needed in original or applied problems.

III. Quantitative Analysis of Operations Involved in Problems: As a further check on these two methods let us bring the third method of determining "relative difficulty" into review: namely, that of making a minute quantitative analysis of the problems in question. This would consist of an enumeration of the like and unlike steps to be carried out by the pupil. To illustrate with the foregoing problems, we arrange the data as in Table III.

The most casual inspection of the data in Table III is enough to lead to the conclusion that the relative difficulty of algebra problems to pupils cannot be determined by a mere quantitative enumeration of the number of like and unlike processes which the pupil must successfully handle. Problem R, a simple substitution problem, involving direct substitution of two quantities and one multiplication, is beyond the abilities of 96.5 per cent of these 169 pupils (presumably a normal group), while Q, a fractional equation including 1 factoring of a trinomial square, 5 multiplica-

tions, 2 additions, 1 subtraction, 1 transposition, 1 changing of sign (11 operations in all), is solved successfully by two-thirds of the group. Clearly "difficulty" is not to be determined in terms of number of operations involved.

TABLE III

Problem	Process	Number of Steps	Substitutions	Total Steps	Percentage of Pupils Solving
D	Addition	1	2	3	67.6
K	Addition	1	2	3	97.0
A	Change signs	1	1		
	Transposition	1		3	91.7
L	Transposition	2	2		
	Change signs	2		6	87.0
O	Squaring	1			
	Multiplication	1			
	Subtraction	1		3	64.5
S	Subtraction	1			
	Squaring	1			
	Multiplication	3	4	9	46.1
T	Addition	7			
	Division	1		8	36.1
P	Transposition	3			
	Change signs	3			
	Subtraction	3		9	53.2
U	Recognition of squaring	2			
	Multiplication	4			
	Addition	1		7	61.0
G	Square root	2			
	Multiplication	1	2	5	34.2
Q	Factoring trinomial square	1			
	Multiplication	5			
	Adding like terms	2			
	Transposition	1			
	Change signs	1			
	Subtraction	1		11	62.8
R	Squaring	1			
	Multiplication	1	2	4	3.5

Thus we repeat, teacher-judgment of the relative difficulty of problems, even when aided by quantitative enumerations of like

and unlike processes, cannot be regarded as valid for the construction of "problem-scales" of definitely evaluated units of difficulty. (From a detailed investigation of the problem in freehand lettering, the writer is also able to state that the teacher-judgment method of determining "merit" in samples of student work—unaided by an objective standard—is to be called into question. This leads him to suspect that results obtained by the methods in drawing and in composition will lead to the same unbalanced results as in algebra.) Obviously we have no satisfactory evidence that the seven problems of Thorndike's scale are separated by equal intervals of difficulty. In fact, from the point of view of the pupil who has to solve the problems, we know that they are not.

This question of the evaluation of the bases upon which educational standards are to be built is fundamental to the whole measuring movement. Many scales have been and are being constructed on the teacher-judgment basis in drawing, composition, handwriting, algebra, and the sciences and applied sciences. Such a slight beginning as we have reported above indicates that the method is questionable. It is recognized that our data are somewhat meager—that a larger group must be used and a much more detailed analysis must be made. The writer is now carrying on an investigation of this question of "bases" as applied to various scales and will report results later. In the meantime we have cleared the way for the organization of our own procedure in establishing standards in algebra. It will be assumed in this preliminary investigation that practicable standard tests can be built on the hypothesis that, in a large group of pupils, relative difficulty will be approximately indicated by the relative proportion of pupils correctly solving the test problems. At the same time it is specifically recognized that the final standardizing of algebra tests must include detailed consideration of a much more complete analysis. This leads us to speak briefly of the fourth method of evaluating "difficulty" in algebra problems, namely:

IV. The Qualitative Psychological Analysis of Problem Solution in connection with carefully controlled tests of individuals and complete introspective and interpretive data by the pupil. As a result of the first seven months of investigation of this problem the writer is convinced that no final standards can be determined

without a thorough utilization of this method. The learning process must be studied both objectively and introspectively with the individual student to reinforce and clarify the data secured from class testing. This latter may result in definite objective standards of the average number of problems "attempted" and "right" in a given unit of time for each specific test given. It will enable the teacher and superintendent to measure their school or system against the norm of many others. But it alone cannot be the sole criterion for the design and construction of the tests themselves.

5. PRINCIPLES GOVERNING THE SELECTION OF TEST PROBLEMS AND CONDUCT OF TESTS

In order that a scale for measuring efficiency in elementary algebra may successfully test automatic efficiency and independent solution it must be composed of two general types of test: (1) a specific test series (A) which will test the specific manipulative abilities of students in all the basic automatic operations involved in the solution of algebra problems; (2) a composite test (B) which will test the independent ability of the student in practical or applied problems.

1. *The Specific Test Series A*, as designed and presented herewith, conforms to the following requirements:

a) It is made up of a series of problem tests each of which is designed as a specific test for a definite automatic operation in algebra solution.

b) Each specific test is made up of a number of problems (10 to 28) each of an elemental nature and involving the operation in question, and each of approximately the same degree of difficulty (estimated here of course).

c) Where it was impossible to arrange separate tests for all kinds of operations involved (owing to lack of time in classroom handling, etc.), those problems which involved closely related operations were grouped in one test and arranged in rotation. Thus the student solving 20 problems may be compared with the one solving 10 problems.

d) Each test was designed as a time test, the time being so arranged (estimated) that no student could quite finish the test in the time given, but so that all could do a considerable number—

otherwise the measure of efficiency would have been too coarse. Care was taken to see that all pupils started and stopped the test at the same instant.

e) The directions were all given orally by the experimenters so that differences in rate of reading and comprehending directions might not complicate results.

f) Test problems were of the alternative sort wherever possible; i.e., they were designed to give either right or wrong answers—otherwise careful evaluation and weighing of answers would have been necessary.

2. *The Composite Test B.*—An ideally designed composite test would be composed of many (say 25) "applied" algebraic problems varying in difficulty by approximately equal intervals and covering all the fundamental types of operations involved in algebraic solution. These problems should not include any of the specific problems of the first test series (A); they should be confined to the abilities of generalization, analysis, and application. Application of the rules and operations for which Test A is the immediate test is the primary function of the composite test. As indicated above, since we cannot use the teacher-judgment method or the quantitative enumeration of operations in this preliminary construction of the composite test, we are forced to the trial-and-error method of selecting a priori what seems to be a representative list of applied problems. The primary purpose of this study has been the standardization of Test Series A and what has been done on Test Series B is of the nature only of a trial-and-error attempt to define the problem and a sound method of approach.

On a basis of the foregoing principles Test Series A (8 specific tests) and Test Series B (composed of 25 applied problems) were constructed. We print below the first few problems of each "specific" test and Test B complete.

THE FIRST SIX PROBLEMS OF THE EIGHT TESTS IN SERIES A

Test I	Test II	Test III
(1) $5(4x-2)=$	(1) $3x+(x+1)=$	(1) $a^3 \cdot a^5=$
(2) $-4(3x-4)=$	(2) $4x+(x-2)=$	(2) $a^x \cdot a^y=$
(3) $-7(2+5x)=$	(3) $n-(-7n+4)n=$	(3) $5a^7 \cdot 6a^5=$
(4) $-3(5-8x)=$	(4) $2y-(y+3)=$	(4) $(n^2)^2=$
(5) $6(-3x-5)=$	(5) $5z-(z-5)=$	(5) $x^{-1}x^1=$
(6) $-8(-4x-7)=$	(6) $3y(y-5)-y^2=$	(6) $(a^6)^{\frac{1}{3}}=$

Test IV. Factoring

(1) $x^2-64 =$
(2) $5x^2+15x=$
(3) $ax^2+bx^2+ay^2+by^2=$
(4) $x^2+4x+4=$
(5) $(x+6)^2-9=$
(6) $x^2-16 =$

Test V

(1) If $a=3$ and $b=2$ what does $a^2 -3ab$ equal?
(2) If $c=6$ and $d=1$ what does $2cd-cd^2$ equal?
(3) If $a=4$ and $b=3$ what does $a^2 -ab^2$ equal?
(4) If $e=5$ and $f=4$ what does $ef -2ef^2$ equal?
(5) If $a=3$ and $b=4$ what does $ab -2ab^2$ equal?
(6) If $x=4$ and $y=6$ what does $2x^2+xy$ equal?

Test VI. Solve for *x*

(1) $-13x=7$
(2) $4x+3=9x - 6$
(3) $7x-5+2x =13$
(4) $4+5(x-3)= 6$
(5) $18x =-31$
(6) $x-4=-8x+14$

Test VII. Solve for *x*
Leave answer in form of fraction

(1) $\frac{4x-2}{3}=\frac{x-3}{4}$
(2) $\frac{-6(2x-8)}{5}+\frac{-3(2x+5)}{4}=\frac{2x+4}{10}$
(3) $\frac{x+1}{x-1}=\frac{5}{3}$
(4) $6x-5-\frac{(6x+11)}{4}=13x$
(5) $\frac{3x-4}{6}=\frac{x+2}{4}$
(6) $\frac{5(x-7)}{7}-\frac{4(3x+6)}{4}=\frac{-4(x-3)}{14}$

Test VIII. Solve for *x*

(1) $x^2-81=0$
(2) $x(x-2)=0$
(3) $x^2+px=6p^2$
(4) $x^2-7x=-12$
(5) $x^2-121=0$
(6) $x(x+7)=0$

ALGEBRA TEST B

1. If a boy is x years old, how old will he be in 5 years? Answer..........
2. An aeroplane that can fly 58.2 miles an hour in still air is retarded by a wind 9.7 miles an hour. At what rate does the aeroplane fly? Answer....................
3. If you represent a number by x, how will you represent 5 more than 4 times the number? Answer......................................
4. Four increased by three times a certain number equals nineteen. Find the number. Answer ....
5. Four years ago a man was seven times as old as his son and his son is now eight years old. Find the age of the father. Answer................
6. A train leaves Pittsburgh for the West at the same time that one leaves for the East. The former travels at the average rate of 42 miles an hour and the latter at the rate of 38 miles an hour. In how many hours will they be 240 miles apart? Answer...............................

7. Which increases more rapidly when r increases, the area of a circle or the circumference? Answer.....................Why?................
8. A line 21 inches long on a certain map corresponds to 22 miles. A line $7\frac{1}{2}$ inches long corresponds to how many miles? Answer...............
9. Eight times a certain number equals 45 diminished by the number. Find the number. Answer..
10. A father is 23 years older than his son and the sum of their ages is 49 years. How old is each? Answer. Father is............Son is.............
11. The perimeter of a rectangle is 256 feet. It is three times as long as it is wide. Find its dimensions. Answer..............................
12. A can do a piece of work in 3 days and B in 4 days. In how many days can both do it working together? Answer........................
13. State whether the quantities mentioned below are directly or inversely proportional:
    a) The number of yards of a certain kind of silk and the total cost. Answer.
    b) Time a train needs to travel 10 miles and speed of train. Answer....
    c) Length of a rectangle of constant width and area of rectangle. Answer...
    d) Distance traveled by train moving at uniform rate and the time. Answer...
14. Given the formula $a=\frac{1}{2}h(b-b')$, find the formula for b'. Evaluate the result for $a=40$; $h=8$; $b=6$. Answer..............................
15. If a boy $4\frac{3}{4}$ feet tall casts a shadow $4\frac{1}{2}$ feet long at the same time that a school building casts a shadow $67\frac{1}{2}$ feet long, how high is the school building? Answer...
16. Find two consecutive numbers whose sum is 157. Answer............
17. If a train moves at the rate of r miles an hour how far will it move in t hours? Answer..
18. The circumference of a circle is given by $2\pi r$; π is $3\frac{1}{7}$ and r is the radius of the circle. If the diameter of a bicycle wheel is 20 inches how many inches in the circumference? Answer.............................
19. A post is $\frac{1}{5}$ of its length in the ground, $\frac{1}{2}$ of its length in water, and 9 feet above water. Find its length. Answer............................
20. Find three consecutive numbers whose sum is 63. Answer............
21. A rectangular field is 10 yards wide and another is 12 yards wide. The second is 5 yards longer than the first and the sum of their areas is equal to 390 square yards. Find the length of each. Answer...............
22. If a boy lying down, with his eye on the ground, sights over the top of a 10-foot pole, held vertically $6\frac{1}{4}$ feet from his eye, and can just see over the top of a tree $37\frac{1}{2}$ feet from his eye, how high is the tree? Answer........
23. A cistern can be filled with two pipes in m and n minutes respectively. In how many minutes can it be filled by the pipes together? Answer....
24. The areas of two circles are proportional to the squares of their radii. If the radii of the two circles are to each other as 4:7 and the area of the smaller circle is 8 square inches, what is the area of the larger? Answer..

25. Find the number whose third and fourth parts added together make 14.
Answer ....

The following teachers kindly co-operated in the investigation by giving two class exercises to the work of testing their pupils and conducting the tests as a regular part of their school work according to our detailed directions: Miss Jessie Brackensiek, Quincy, Ill.; Dr. E. H. Taylor and his teachers of mathematics in the State Normal School, Charleston, Ill.; Mr. Ward Taylor and his teachers of mathematics in the State Normal School, Carbondale, Ill.; Miss Lida C. Martin and Mrs. Hostetler, Decatur, Ill.; Miss Fannie Andrews, Marshall, Ill.; Mr. R. M. Ginnings, Macomb, Ill.; Mr. O. R. Hedden, Robinson, Ill.; Mr. E. L. Moyer, Chicago Heights, Ill. Our thanks are due to these teachers for their co-operation. It should be stated that at least fifteen teachers in other systems expressed their hearty interest in the problem but were prevented from co-operating only by the fact that their final examinations for the year were then being held. The reorganization of our investigation, coming as it did in the spring, necessitated this late request for assistance—otherwise a much larger amount of data would have been gathered. We have had, however, all the data that there was time to consider and more would have delayed the sending out of this report.

Complete tests were taken by 518 pupils who were just finishing first-year algebra, giving us something over 4,500 test sheets in all. The treatment of these data may be outlined as follows:

(1) Original tabulations: The tests were next corrected and tabulated in terms of the number of problems "attempted" and "right" for each pupil.

(2) The average number of problems attempted and worked correctly in each test for each school, in one minute, and the averages for the entire group of pupils were next computed, *by finding the harmonic mean.*[1] For purposes of this preliminary

[1] In this connection the reader's attention is called to the fact that in averaging data involving "time rates" it is necessary to use the *harmonic mean*, which is the arithmetic mean of the sum of the reciprocals of all the numbers. A statistical fallacy common to educational investigations involving time rates (e.g., see Starch, *The Measurement of Efficiency in Reading, Writing, and Spelling*, Madison, Wis., 1915) is to average such rates by the arithmetic mean. This practice invariably gives a

investigation these averages may be regarded as tentative standards against which to check the efficiency of the teaching process in any school.

(3) The rank of each school was next computed for each test and the average rank of each school in all tests combined.

(4) The relationship existing between the abilities involved in the different tests was computed by the Pearson product-moment method.

(5) The percentage of problems attempted which were worked correctly in each test, i.e., the relationship between speed and accuracy in various types of test, was worked out.

(6) The tabulation of particular problems correct in each test and a study of the validity of the method of constructing Test A was next taken up.

(7) A detailed tabulation of the particular errors made by each of 100 pupils (selected at random) in each test.

(8) A program of procedure for the continuation of the study.

In order to accommodate the data of this study to the space of a magazine article they will be presented in tabular form with brief interpretation and discussion of each table.

Table IV presents the typical features of our data for eight school systems. These data provide us with tentative standards only concerning the number of problems (consisting of definite types of operations) which should be worked in a given unit of time. Given a large and representative sampling of schools, we shall be able to refine our standard tests and their applicability to practical school measurement. For the time being the primary value of our results lies in the fact that they provide the basis for a critical revision of our tests, indicate the general line of approach in this problem of test construction, and lay bare certain weaknesses in the standardizing of the "teaching emphasis." These problems are fundamental, for the question of principles-of-design and

result deviating from the true mean by as much as 15 per cent. The use of the arithmetic mean in averaging the time rates of this investigation results in an error exceeding that amount. Proof and illustrations of the effect of using the two means will be ound in the first of a series of discussions of "Statistical Fallacies in Educational Research," shortly to be published elsewhere by the present writer.

methods-of-construction must soon be worked out to a satisfactory conclusion.

TABLE IV

THE AVERAGE NUMBER OF PROBLEMS PER MINUTE "ATTEMPTED" AND "RIGHT" FOR EACH OF EIGHT SCHOOLS AND FOR ENTIRE GROUP OF 518 PUPILS

Tests ...	I		II		III		IV		V		VI		VII		VIII		No. Minutes per Problem Test B	
School ..	Att.	Rt.	Att.	Rt.	Att.	Rt.	Att.	Rt.	Att.	Rt.	Att.	Rt.	Att.	Rt.	Att.	Rt.	Att.	Rt.
A	8.34	6.60	3.45	1.38	7.50	4.30	2.17	1.38	2.17	1.27	2.85	1.04	0.69	0.17	0.76	0.37	2.45	3.63
B	7.95	5.35	3.25	1.08	7.19	4.00	2.70	1.60	2.56	1.08	2.30	0.92	.57	.19	1.38	.65	4.00	6.20
C	10.06	8.00	3.37	1.15	8.57	4.72	1.57	0.79	2.71	0.98	1.68	0.63	.56	.17	1.23	.33	2.65	4.65
D	8.39	7.89	3.41	0.79	5.00	2.30	2.24	0.84	2.38	1.04	1.60	0.97	.95	.26	1.25	.58	....	....
E	9.37	9.16	4.37	2.11	4.95	3.05	1.85	1.08	2.44	1.31	2.70	1.28	.58	.22	0.87	.41	2.50	3.65
F	9.91	8.36	4.48	2.05	7.45	3.85	2.70	2.25	2.80	1.63	3.17	1.38	.83	.26	1.45	.54	2.97	4.36
G	6.54	4.45	4.41	1.24	6.89	2.59	2.21	1.31	2.63	0.96	2.70	1.25	.72	.26	....	....	2.45	5.06
H	10.80	7.67	3.85	0.95	8.39	2.67	2.41	0.92	2.72	1.14	3.38	1.35	0.85	0.21	0.99	0.41	3.19	5.45
Average of schools	8.30	6.36	3.92	1.29	7.06	3.49	2.26	1.19	2.57	1.11	2.50	1.09	0.75	0.23	1.11	0.51	2.99	4.79
Highest..	10.06	9.16	4.48	2.11	8.57	4.72	2.70	2.25	2.80	1.63	3.38	1.38	0.95	0.26	1.45	0.65	4.00	6.20
Lowest..	6.54	4.45	3.25	0.79	4.95	2.30	1.57	0.79	2.17	0.96	1.60	0.63	0.56	0.17	0.76	0.33	2.45	3.63

Several outstanding facts may be set down:

(1) Pupils can correctly solve five times as many problems of the type of Test I (simple removal of parentheses and changing of signs) as they can of Tests II, IV, V, VI—this in spite of the fact, for example, that the average number and kinds of steps and operations necessary for solving each problem in the first two tests are almost the same, being for Test I, 3.5, and for Test II, 4.2.

(2) About the same number of problems can be solved per minute in Tests II, IV, V, and VI, again regardless of the kinds or number of operations used in each test. These data reinforce our conclusions that the relative difficulty of different types of problems cannot be obtained by a quantitative analysis of the problems concerned. Again, we have to date no known method of equating difficulty in different types of algebra problems. These points emphasize our contention that any standard test should be so designed as to test for one or at most two or three fundamental operations.

Furthermore, the data given above bear directly on the question, To what extent is instruction in first-year algebra making habitual

certain "mechanical" processes represented by these tests? The answer is direct: To only a very slight extent. Instruction resulting in a capacity for solving correctly less than one factoring problem per minute cannot be said to be efficient, especially when one of the eight schools is securing a skill three times as great. *It is the higher attainments represented by the work of these schools that are of significance to us, rather than the types, as they point out the possibility of practicable efficiency.* This should be the primary function of the standard test.

The blame cannot be laid solely to the door of classroom instruction. Rather it should be laid on the teacher's *organization* of her classwork as expressed through her *teaching emphasis*. The figures given above show that no school is highly efficient or decidedly deficient in perfecting all kinds of processes. All schools are relatively efficient in the perfecting of some of these habitual processes. Clearly our results point to a great variability of teaching emphasis, some teachers spending their time making automatic certain operations, others laying the stress on other processes. Our data point to the need of a statement of the standard amount of efficiency to be worked for by each teacher of algebra and of a more intensive study of methods of perfecting the particular manipulations.

A seeming lack in the presentation of the foregoing data is some measure of variability. This will be expressed in the final standards in the form of the limits of the middle 50 per cent of the distribution. It has not been deemed expedient to take the time to compute these in this preliminary report.

Our data thus point to lax methods of habit formation in the learning of these mechanical processes and to a decided lack of emphasis on ideals or attitudes of accuracy. But the most pertinent evidence which we have on these points is found in connection with an analysis of (1) the particular problems which pupils are able to solve in each test, and (2) the specific errors made by pupils in solving particular problems.

It has been assumed that difficulty-to-the-student will be indicated approximately by the percentage of pupils correctly solving the problems in the tests. Careful study of the data in Table VI

will therefore aid us materially in the redesigning of our standard tests. It was intended (although not exactly carried out) to design the tests in accordance with the principle that the problems should be arranged in cycles, those problems necessitating exactly the same steps recurring every so many problems. This would enable us to compare the record of one pupil with that of another without having a specific test for every particular operation used. The data shown in Table VI for Tests I, III, IV, V, and VII in the main conform to this practice. That is, those problems utilizing the same specific operations are, in general, correctly solved by approximately the same proportion of pupils. In the remaining tests, II, VI, VIII, however, while there is a rough paralleling of types-of-operations and proportion-of-pupils-solving, there is need for a much more detailed study of the problem. As a definite guide to the standardizing of the tests the data will be of great service. The study of the returns to date seem to indicate that the cycle principle must be adhered to strictly throughout the tests. The recurring problems must involve exactly the same steps in solution and the cycles must be of exactly the same size—otherwise the material will not offer comparable results.

The standardizing of tests and of the teaching process will, furthermore, be much clarified by a detailed study of the particular errors which pupils make in solving these problems. Table VI and the descriptive list following present the data on this point.

DESCRIPTION OF ERRORS ACCORDING TO NUMBER IN EACH TEST

Test I

No. of Error in Table VII

1. Incorrect multiplication of figures.
2 Failure to change signs with − sign outside parenthesis.
3. Failure to use correctly with − sign inside parenthesis.
4. Failure to follow directions.
5. Error in writing.
6. Used + sign inside parenthesis incorrectly.

Test II

1. Incorrect multiplication of figures.
2. Failure to change signs with − sign outside parenthesis.
3. Failure to use correctly with − sign inside parenthesis.

TABLE V

THE PERCENTAGE OF 100 PUPILS (SELECTED AT RANDOM WITH AN EQUAL PROPORTION FROM EACH SCHOOL REPRESENTED) WHO WORKED CORRECTLY CERTAIN PROBLEMS IN TESTS I TO VIII AND TEST B

Problems	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Test I	96	88	82	87	91	84	82	94	99	85	79	98	...	...	...	...	...	...	...	...	...	...	...	...	...
" II	57	56	29	44	55	66	54	73	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
" III	91	50	84	98	14	63	75	90	37	86	85	19	39	...	...	...	...	...	...	...	...	...	...	...	...
" IV	83	38	32	85	25	78	41	30	74	30	85	34	42	89	49	...	...	...	...	...	...	...	...	...	...
" V	55	54	51	54	57	53	57	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
" VI	69	64	66	47	75	64	51	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
" VII	31	...	35	8	46	2	57	0	39	0	43	9	...	...	...	...	...	...	...	...	...	...	...	...	...
" VIII	91	32	24	51	85	27	51	22	44	37	36	41	68	...	...	...	...	...	...	...	...	...	...	...	...
" B	68	54	89	80	46	38	31	13	61	67	42	28	20	6	12	61	66	16	37	79	33	0	16	8	70

TABLE VI

TYPES AND NUMBER OF ERRORS MADE BY 100 PUPILS IN WORKING PARTICULAR PROBLEMS IN TESTS I, II, III, V, AND VI

Number of Error	Problem Number																												Total Number	Percentage of Total	No. per Problem
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28			
													Test I																		
1	3	2	3	4	2	8	1	1	1	4	15	1	...	2	2	1	2	2	2	3	1	3	4	...	...	...	...	...	67	28.2	3.0
2	...	10	16	9	...	10	17	...	...	...	8	...	...	...	9	...	9	...	10	1	...	7	8	...	...	...	...	...	114	47.9	9.5
3	...	...	...	...	7	...	...	5	...	9	...	...	7	...	...	1	...	6	...	...	2	...	...	2	...	...	...	...	39	16.4	4.9
4	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	2	0.8	1.0
5	...	...	...	...	...	...	...	...	...	2	3	1	2	...	2	...	...	...	2	1	...	1	...	...	...	...	...	...	14	5.9	1.8
6	...	...	...	...	...	...	...	...	...	...	...	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	2	0.8	1.0
Totals	4	13	19	13	9	18	18	6	1	15	26	3	10	2	13	2	11	8	14	5	3	11	12	2	...	...	...	...	238		
														Test II																	
1	1	2	2	3	2	9	11	7	1	5	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	44	7.6	4.0
2	...	...	29	29	19	1	25	9	16	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	128	22.2	18.3
3	...	...	...	...	...	...	...	...	...	...	...	...	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1	0.2	1.0
4	1	1	2	6	3	10	5	4	3	5	3	2	1	2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	48	8.3	3.4
5	...	1	2	...	2	1	3	1	2	...	...	...	...	5	...	...	...	...	...	...	...	...	...	...	...	...	...	...	17	3.0	2.1
6	33	35	27	25	23	4	14	12	8	1	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	183	31.8	16.6
7	7	5	9	4	4	5	6	4	...	2	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	48	8.3	4.4
8	...	...	...	...	1	1	2	2	...	1	...	3	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	11	2.0	1.6
9	1	1	40	...	1	1	...	...	...	1	...	4	3	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	52	9.0	6.5
10	...	...	...	2	1	2	1	1	...	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	8	1.4	1.3
11	...	...	2	...	1	5	...	1	...	2	...	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	13	2.3	1.9
12	1	2	3	...	2	4	...	...	1	...	1	2	...	2	...	...	...	...	...	...	...	...	...	...	...	...	...	...	18	3.1	2.0
13	1	1	1	1	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	6	1.0	1.0
Totals	45	48	117	70	60	44	67	41	31	18	5	15	6	10	...	...	...	...	...	...	...	...	...	...	...	...	...	...	577		

														TEST III																
1		2	7							6					3		13				2					1		1	35	4.3	4.4
2	4	53	5					7	51	9		2			8	55						7		30					231	28.5	21.0
3						7		1			14		8	1				21		9					14				75	9.3	9.4
4							2														3							2	7	0.9	2.3
5			1			3											3				2								9	1.1	2.3
6														1															1	0.1	1.0
7						1							1							8	3					8			21	2.6	4.2
8																	74				2		21			1			98	12.1	32.7
9					50																					1			51	6.3	25.5
10												29																	29	3.6	29.0
11																											24		24	3.0	24.0
12													1													2			3	0.4	1.5
13	4	11	2		5	8	12	2	10	1		20	19	4	2	8	7	5	19	5	19	4	3	9	2	6	6	3	196	24.2	7.5
14																		5							3				8	1.0	4.0
15			1	1	2	1	1					2														2			10	1.2	1.4
16															1								1	1	1				4	0.5	1.0
17																		1											1	0.1	1.0
18													3																3	0.4	3.0
29						3																							3	0.4	3.0
Totals...	8	66	16	1	57	23	15	10	61	16	14	53	32	6	14	63	97	32	19	22	31	11	25	40	20	21	30	6	809		

TABLE VI—*Continued*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total Number	Percentage of Total	No. per Problem
									TEST V									
1...	14	10	13	14	11	8	5	5								80	22.1	10.0
2...	6	5	19	7	7											44	12.2	8.8
3...	14	11	11	13	7	6	4	6								72	19.9	9.0
4...																		
5...			1		3											4	1.1	2.0
6...		1							1							2	1.6	1.0
7...	10	11	8	9	10	9	5	3	1							66	18.2	7.3
8...	10	9	8	11	6	3	1	2								50	13.8	6.3
9...	2	4	5	4	8		1									24	6.6	4.0
10...	1	2	2	5	3	2		1	1							17	4.7	2.1
11...							1									1	0.3	1.0
12...																		
13...	1					1		1								3	0.8	1.0
Totals	58	53	67	63	55	29	17	18	3	...	...	...	...	...	...	363	....	....

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total Number	Percentage of Total	No. per Problem
									TEST VI									
1...																		
2...	5	34	9	8	10	15	24	10	10	12	6	2	1			146	39.5	11.2
3...																		
4...																		
5...				1	3		1									5	1.4	1.7
6...		6	7	21	1	6	10	4		3	3					61	16.5	6.8
7...																		
8...																		
9...		3	1	6	4	4	4	7	2	2	6	4	1	1	1	46	12.5	3.3
10...							2			1						3	0.8	1.5
11...			9						1							10	2.5	5.0
12...	1															1	.3	1.0
13...	6	10	9	8	6	7	7	6	3	5	3	2	1			73	19.7	5.6
14...		2		2												4	1.1	2.0
15...		5	1	3	3	4	1		2	1	1	1				22	6.0	2.2
.....	12	50	36	49	27	36	49	27	18	24	19	9	3	1	1	371	....	....

4. Failure to combine like terms or reduce to simplest form.
5. Error in writing; omission of symbols, etc.
6. Ignoring + or − sign and multiplying across them.
7. Ignorance of operations needed.
8. Addition instead of multiplication symbol following parenthesis.
9. Omitted multiplication by symbol preceding or following parenthesis.
10. Incorrect addition or subtraction of like terms.
11. Omitted terms in the answer.
12. Faulty multiplication of symbols.
13. Failure to follow directions.

TEST III

1. Incorrect multiplication of figures.
2. Multiplied exponents instead of adding them.
3. Added exponents instead of multiplying them.
4. Divided exponents instead of subtracting them.
5. Error in writing.
6. Multiplied exponents instead of subtracting them.
7. Multiplied denominator instead of numerator.
8. Incorrect addition of exponents, e.g., $9x^m \cdot 7x = 63x^m$ instead of $63x^{m+1}$.
9. Regarding x^{-1} as 1 instead of $\frac{1}{x}$.
10. Regarding $y^{-2}x^2$ as xy (second powers canceling each other).
11. $(x^0)^2$ called x^2; 0 power equivalent to first power.
12. Subtracting exponents instead of multiplying them.
13. General inability to use operation.

14–19. Scattering errors.

TEST V

1. Incorrect multiplication of figures.
2. Incorrect addition or subtraction (− quantity the larger).
3. Substitution of wrong figures.
5. Error in writing.
6. Incorrect addition or subtraction (+ quantity the larger).
7. Incomplete.
8. Inability to solve this type of problem.
9. Incorrect substitution.
10. Addition in place of multiplication.

11–13. Scattering errors.

TEST VI

2. In transposing terms neglected to change signs.
5. Error in writing.
6. Incorrect addition.
9. Incomplete, not reduced to fractional form.
11. Subtracted instead of added.
13. Inability to handle operation.
14. Multiplied across + or − signs before parenthesis.
15. Incorrect division (inverted numerator and denominator).

Discussion of specific types of error—Test I: Over one-fourth of the errors made in this simple test may be ascribed to lack of fixing the multiplication tables in arithmetic. It is necessary to recall that Test I shows evidence of being fairly well automatized—the only test in the eight of which this can be said. Furthermore there being but few different algebraic processes involved (only two to five operations possible) the actual proportion of total errors found to be due to lack of automatization of fundamentals in arithmetic must necessarily be larger. The actual number of errors of any one type is small. Even so, each of the pupils on the average made two errors in solving twelve problems of this simple type. It is clear, however, that there is a need for more intensive drill on the process of removing parentheses, especially of the type of problem having the minus sign before the parenthesis. Table VII shows that the problems involving this operation, 6, 7, 11, 19, etc., are the ones in which errors most frequently occur and that the problem is much more difficult to the student if a minus sign is used both inside and outside the parenthesis. The data confirm the view that *teachers should analyze problems into constituent operations and drill on those that prove most difficult.* The use of the minus sign is the stumbling-block in this most simple of algebraic operations.

Test II: With a but slightly more complex problem and twice as many possible types of error we find in Test II a relatively small proportion of arithmetic errors. In 100 pupils there were found on the average four who made multiplication mistakes in solving each problem. The results of Test II confirm our interpretation of Test I that *the use of the minus sign in removing parentheses is a source of weakness (22.2 per cent of errors were of type 2)* and an element of instruction on which greater stress should be laid in drill. One-third of the nearly 600 errors made by 100 pupils in solving twelve problems are made in ignoring a plus or minus sign before parentheses (error No. 6). The data show that this mistake is common to all of the eight schools but one. Clearly this is a type of error that should have been eliminated by class drill. To repeat: First-year algebra should make automatic the removal of parentheses. The evidence leads us to think that it is not doing so. If the data presented here are valid, 8 per cent of the 100 pupils

show distinct evidence of absolute inability to handle this operation. About 15 per cent of the errors may be ascribed specifically to carelessness in writing (omitting terms, etc.). Nearly all the errors that were made emphasize the careless and slipshod work done by our pupils, a condition due primarily to weaknesses in the teaching process.

Test III: In the manipulation of exponents a few definite types of errors occur, viz.: Exponents are multiplied when they should be added and vice versa; problems of the type $9x^m \cdot 7x = ?$ are almost invariably solved as $63x^m$ (error No. 8; another example of multiplying exponents instead of adding them?); the use of negative exponents has not been permanently fixed by at least half the pupils (errors Nos. 9 and 10); the use of the zero power has not been learned; 24 per cent of all errors made point to an inability to handle exponents in general. *The tabulations above emphasize again the need for analysis and sound distribution of teaching emphasis on the difficult operations.* Complete mastery of the fundamental operations should be insisted upon.

Test V: The surprisingly large number of errors made in this test bring out again the weakness in the mastery of the arithmetic fundamentals: on the average over three errors per person in solving eight problems. Furthermore, one pupil in five made an error in addition, subtraction, or multiplication in solving eight problems. Further evidence of careless methods is seen in the fact that 20 per cent of the errors made are errors in substituting wrong numbers and 20 per cent more are due to incomplete solution. As with the preceding relatively simple operations so with simple substitution, many pupils are not able to handle successfully this process. *Clearly, if elementary algebra should mechanize any "preparatory" operations that are used repeatedly in problems requiring independent thinking it should mechanize simple substitution. That it is not doing so is strongly suggested by our data.*

Test VI: In the solution of simple equations of the first degree with one unknown, eleven pupils out of a hundred neglect to change signs in transposing terms. Two-fifths of all errors made are of this type. The deduction from such data is clear: *a primary purpose of elementary algebra should be to make automatically accurate*

just such processes as changing signs in transposing terms. Teachers interested in improving the outcomes of instruction should use some such method for determining the difficult elements in the "learning process" in algebra, and redistribute their teaching emphasis in conformity to the difficulty of the operations in question. It is just such analyses as we are making here that will suggest to the teacher "relative difficulty" of operations and will point to a sound solution of the problem.

SUMMARY OF CONCLUSIONS

1. The subject-matter of first-year algebra should be definitely organized in the form of a specific statement of (*a*) the "mechanical" processes which should be drilled until perfectly habitualized; (*b*) the typical "original" or applied problems in which should be given at least a definite minimum of practice in the application of the mechanical processes to new problematic situations. (Such an organization is tentatively outlined above.)

2. The efficiency of instruction in the mechanical processes may be tested by a series of time tests of the nature of those given herewith, each test being so designed as to test for one, or at most two or three, closely related processes.

3. The efficiency of instruction in developing skill in the solution of "original" problems may be measured by a standard scale of representative problems (the relative difficulty of each problem having been determined by some of the methods discussed herewith), the problems being separated by definitely determined intervals of difficulty.

4. The study leads to the conclusion that the relative difficulty of the problems composing standard tests cannot be determined by the teacher-judgment method. There is evidence to support the view that difficulty-to-the-pupil will be indicated, at least approximately, by the proportion of a large group of pupils solving the problems in question. Difficulty of problems and relative "teaching emphasis" can be determined completely only by a detailed psychological analysis of each of the mental processes involved in the learning of algebra.

5. Standard tests, if representative of the conditions of algebra instruction generally, may be of definite service to the teacher and the supervisory officer by providing (1) objective measurement of the efficiency of instruction; (2) means of comparing efficiency with that of other representative teachers and systems; (3) means of determining particular weaknesses in the learning and teaching processes.

6. The study of errors made by pupils indicates that *inefficiency in algebraic solution is due primarily to lack of mastery (habitualization) of a few typical operations which recur frequently in such solution.* (Such operations include, e.g., the use of the minus sign in removing parentheses, principles concerning the addition and multiplication of exponents, the use of negative exponents, the use of the zero power, simple substitution, neglecting to change signs in transposing terms; etc.) This condition points to a need for a thorough study of (1) the psychology of the learning process in algebra; (2) the relative emphasis that should be placed on the teaching of certain processes, i.e., the relative drill emphasis.